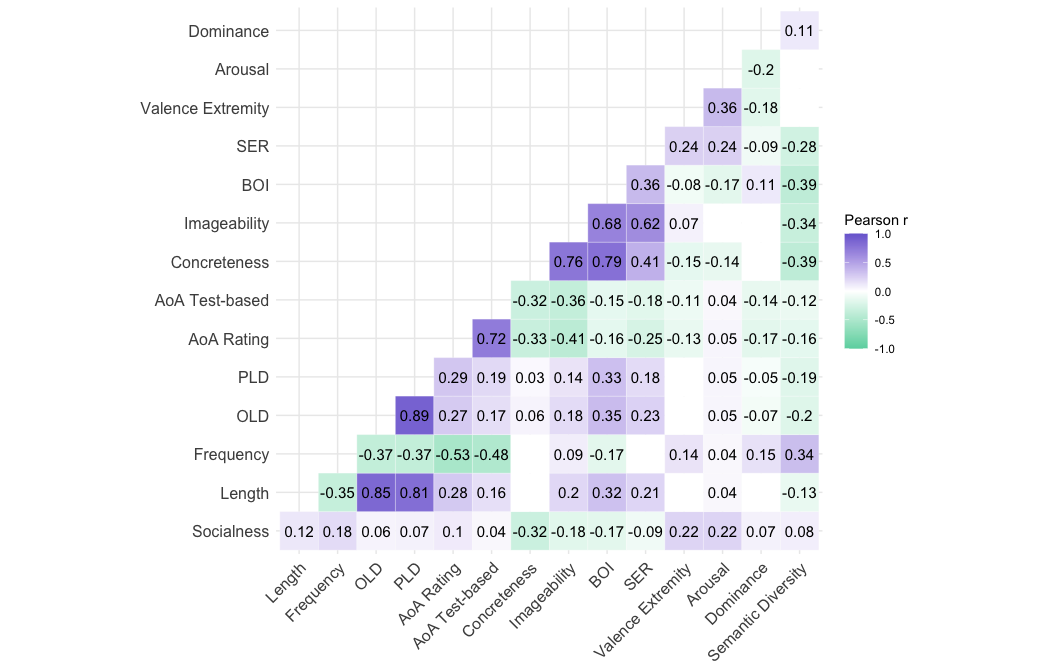
Socialness Ratings: Analyses

Veronica Diveica



**Figure 2**Correlations between socialness ratings and lexical-semantic dimensions. Only correlations significant at *p* < .01 are shown. The strength and direction of the correlation coefficients are indicated by the colour and the numerical values. For each variable of interest, the numbers of items in common with our socialness ratings are as follows: valence, arousal and dominance: 8,388; length: 8,388; log subtitle frequency: 8,160; OLD: 8,027; PLD: 8,027; rating-based AoA: 8,348; test-based AoA: 7, 321; concreteness: 8,388; imageability: 2,680; BOI: 4,038; SER: 2,645. SER = sensory experience rating; BOI = body-object interaction; AoA = age of acquisition; PLD = phonologic Levenshtein distance; OLD = orthographic Levenshtein distance.

**Correlations with Lexical and Semantic Properties**

We examined the correlations between the socialness ratings and various lexical and semantic properties of the words. The lexical variables included letter length, orthographic Levenshtein distance (Yarkoni et al., 2008), phonological Levenshtein distance and frequency (log subtitle frequency; Brysbaert & New, 2009) and assessed whether the socialness measure captures basic lexical attributes. To assess the relationship between the socialness ratings and sensorimotor information, the semantic variables included concreteness (the degree to which the word’s referent can be experienced through one of the five senses ; Brysbaert et al., 2014), imageability (the ease with which the word arouses a mental image ; Cortese & Fugett, 2004; Schock et al., 2012), body-object interaction (BOI; the ease with which a human body can physically interact with a word’s referent; Pexman et al., 2019), and sensory experience ratings (the degree of sensory experience evoked; Juhasz & Yap, 2012). To assess the relationship between the socialness ratings and affective information, the semantic variables included valence extremity (i.e. the absolute value of the difference between the valence rating from 5, the neutral point on the scale measuring the degree to which the word evokes positive as opposed to negative feelings; Warriner et al., 2013), arousal (the degree to which the word evokes feelings of arousal as opposed to calm; Warriner et al., 2013), and dominance (the degree to which the word evokes feelings of being controlled as opposed to in control; Warriner et al., 2013). Finally, to assess the relationship between the socialness ratings and linguistic experience, the semantic variables included semantic diversity (the extent to which a word appears in semantically-diverse contexts; Hoffman et al., 2013), rating-based age of acquisition (AoA) (Kuperman et al., 2012), and a test-based AoA measure derived from (Dale & O’Rourke, 1981) and updated by (Brysbaert & Biemiller, 2017).

These correlations revealed several interesting relationships that provide insight as to the nature of the word socialness measure (Figure 2; see Figure S1 for scatterplots). Socialness negatively correlated with concreteness (*r* = -0.32), imageability (*r* = -0.18), and BOI (*r* = -0.17), which suggests that words with less social relevance are associated with more embodied sensorimotor information. In contrast, socialness ratings positively correlated with valence extremity (*r* = 0.22), and arousal (*r* = 0.22), suggesting that social words tend to have more affective information.

(#tab:LDT regressions tables)

*Means, standard deviations and correlations of all variables of interest for the regression analysis predicting performance in the English Lexicon Project Lexical Decision Task (N = 6,926)*

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Variable | M | SD | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| 1. Length | 7.54 | 2.13 |  |  |  |  |  |  |  |  |
| 2. Frequency | 2.18 | 0.65 | -.36\*\* |  |  |  |  |  |  |  |
| 3. Age of Acquisition | 9.43 | 2.44 | .31\*\* | -.57\*\* |  |  |  |  |  |  |
| 4. Socialness | 3.67 | 1.24 | .13\*\* | .18\*\* | .09\*\* |  |  |  |  |  |
| 5. Concreteness | 3.08 | 0.95 | -.06\*\* | .07\*\* | -.35\*\* | -.29\*\* |  |  |  |  |
| 6. Valence Extremity | 1.07 | 0.77 | .01 | .14\*\* | -.15\*\* | .23\*\* | -.14\*\* |  |  |  |
| 7. Semantic Diversity | 1.60 | 0.31 | -.13\*\* | .34\*\* | -.16\*\* | .09\*\* | -.40\*\* | .03\* |  |  |
| 8. LDT zRT | -0.25 | 0.31 | .52\*\* | -.59\*\* | .56\*\* | -.03\* | -.11\*\* | -.08\*\* | -.25\*\* |  |
| 9. LDT Error Rate | 0.06 | 0.08 | -.08\*\* | -.33\*\* | .37\*\* | -.10\*\* | -.06\*\* | -.10\*\* | -.15\*\* | .51\*\* |

*Note.* M and SD are used to represent mean and standard deviation, respectively. LDT = lexical decision task; zRT = standardized reaction times. \* indicates p < .05. \*\* indicates p < .01.

(#tab:LDT regressions tables)

*Regression Coefficients from Item-Level Analyses Predicting Lexical Decision Task Latencies and Accuracy (N = 6,926)*

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Predictor | b | SE | t | p | sr2 | R2 | ∆R2 | b | SE | t | p | sr2 | R2 | ∆R2 |
| Step1 |  |  |  |  |  | 0.51 |  |  |  |  |  |  | 0.21 |  |
| Intercept | -0.246 | 0.003 | -94.49 | 0 |  |  |  | 0.063 | 0.001 | 70.97 | 0 |  |  |  |
| Length | 0.047 | 0.001 | 35.6 | 0 | 0.09 |  |  | -0.01 | 0 | -22.57 | 0 | 0.058 |  |  |
| Frequency | -0.15 | 0.005 | -29.99 | 0 | 0.064 |  |  | -0.033 | 0.002 | -19 | 0 | 0.041 |  |  |
| Age of Acquisition | 0.035 | 0.001 | 26.91 | 0 | 0.051 |  |  | 0.01 | 0 | 22.99 | 0 | 0.06 |  |  |
| Step2 |  |  |  |  |  | 0.52 | 0.006 |  |  |  |  |  | 0.22 | 0.006 |
| Intercept | -0.246 | 0.003 | -95.06 | 0 |  |  |  | 0.063 | 0.001 | 71.21 | 0 |  |  |  |
| Length | 0.048 | 0.001 | 35.75 | 0 | 0.089 |  |  | -0.01 | 0 | -21.5 | 0 | 0.052 |  |  |
| Frequency | -0.13 | 0.005 | -23.9 | 0 | 0.04 |  |  | -0.028 | 0.002 | -14.7 | 0 | 0.024 |  |  |
| Age of Acquisition | 0.038 | 0.001 | 25.78 | 0 | 0.046 |  |  | 0.011 | 0.001 | 22.31 | 0 | 0.056 |  |  |
| Socialness | -0.011 | 0.002 | -4.73 | 0 | 0.002 |  |  | -0.003 | 0.001 | -3.57 | 0 | 0.001 |  |  |
| Concreteness | 0 | 0.004 | 0.02 | 0.984 | 0 |  |  | 0.002 | 0.001 | 1.7 | 0.088 | 0 |  |  |
| Valence Extremity | 0.007 | 0.004 | 1.83 | 0.067 | 0 |  |  | -0.001 | 0.001 | -0.64 | 0.525 | 0 |  |  |
| Semantic Diversity | -0.068 | 0.01 | -6.77 | 0 | 0.003 |  |  | -0.012 | 0.003 | -3.54 | 0 | 0.001 |  |  |

*Note.* b represents unstandardized regression weights. SE represents the standard error of the regression weights. sr2 represents the semi-partial correlation squared. LDT lexical decision task. zRTs standardized reaction times. *p < .05;* ***p < .01;*** p < .001

There were 6,926 items for which we had values for all variables of interest in the analysis predicting LDT performance. Descriptive statistics and zero-order correlations between all variables of interest from this dataset are reported in supplementary Table S1. The statistical results are reported in Table 3 and the standardized coefficients are illustrated in Figure 3A. In this analysis, the control variables were all significant predictors of LDT latencies – RTs were faster for words that are shorter, more frequent and acquired earlier. There was significant improvement in model fit with the addition of the semantic variables, which collectively accounted for a further 0.61% of variance in LDT latencies. Of the semantic variables, only socialness and semantic diversity were significant predictors, with faster RTs for words with increased social relevance and those encountered in more semantically diverse contexts. A similar pattern of results was observed when predicting LDT error rates. The control variables were all significant predictors, with less errors for words that are longer, more frequent and acquired earlier. There was significant improvement in model fit with the inclusion of the semantic variables, which accounted for an additional 0.56% of variance in LDT error rates. Socialness and semantic diversity were the only significant semantic predictors – error rates were lower for words with increased socialness and for those that are more semantically-diverse.

(#tab:ECP regressions tables )

*Means, standard deviations and correlations of all variables of interest for the regression analysis predicting performance in the English Crowdsourcing Project Word Knowledge Task (N = 7,010)*

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Variable | M | SD | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| 1. Length | 7.55 | 2.13 |  |  |  |  |  |  |  |  |
| 2. Frequency | 2.17 | 0.65 | -.36\*\* |  |  |  |  |  |  |  |
| 3. Age of Acquisition | 9.43 | 2.43 | .31\*\* | -.57\*\* |  |  |  |  |  |  |
| 4. Socialness | 3.67 | 1.24 | .13\*\* | .18\*\* | .09\*\* |  |  |  |  |  |
| 5. Concreteness | 3.08 | 0.95 | -.06\*\* | .07\*\* | -.35\*\* | -.29\*\* |  |  |  |  |
| 6. Valence Extremity | 1.07 | 0.77 | .01 | .14\*\* | -.14\*\* | .23\*\* | -.14\*\* |  |  |  |
| 7. Semantic Diversity | 1.59 | 0.31 | -.13\*\* | .34\*\* | -.16\*\* | .09\*\* | -.40\*\* | .03\* |  |  |
| 8. Recognition zRT | -0.53 | 0.12 | .39\*\* | -.55\*\* | .53\*\* | -.05\*\* | -.11\*\* | -.14\*\* | -.23\*\* |  |
| 9. Proportion Unknown | 0.01 | 0.02 | -.06\*\* | -.35\*\* | .38\*\* | -.07\*\* | -.06\*\* | -.13\*\* | -.18\*\* | .63\*\* |

*Note.* M and SD are used to represent mean and standard deviation, respectively. zRT = standardized reaction times. \* indicates p < .05. \*\* indicates p < .01.

(#tab:ECP regressions tables )

*Regression Coefficients from Item-Level Analyses Predicting ECP Word Knowledge Task Latencies and Accuracy (N = 7,010)*

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Predictor | b | SE | t | p | sr2 | R2 | ∆R2 | b | SE | t | p | sr2 | R2 | ∆R2 |
| Step1 |  |  |  |  |  | 0.4 |  |  |  |  |  |  | 0.23 |  |
| Intercept | -0.527 | 0.001 | -495.33 | 0 |  |  |  | 0.013 | 0 | 69.57 | 0 |  |  |  |
| Length | 0.01 | 0.001 | 19.37 | 0 | 0.032 |  |  | -0.002 | 0 | -22.24 | 0 | 0.055 |  |  |
| Frequency | -0.056 | 0.002 | -27.65 | 0 | 0.065 |  |  | -0.007 | 0 | -19.99 | 0 | 0.044 |  |  |
| Age of Acquisition | 0.014 | 0.001 | 25.28 | 0 | 0.054 |  |  | 0.002 | 0 | 24.02 | 0 | 0.064 |  |  |
| Step2 |  |  |  |  |  | 0.41 | 0.008 |  |  |  |  |  | 0.23 | 0.008 |
| Intercept | -0.527 | 0.001 | -498.44 | 0 |  |  |  | 0.013 | 0 | 69.93 | 0 |  |  |  |
| Length | 0.011 | 0.001 | 20.21 | 0 | 0.034 |  |  | -0.002 | 0 | -21.67 | 0 | 0.051 |  |  |
| Frequency | -0.049 | 0.002 | -22.07 | 0 | 0.041 |  |  | -0.006 | 0 | -15.74 | 0 | 0.027 |  |  |
| Age of Acquisition | 0.013 | 0.001 | 22.5 | 0 | 0.043 |  |  | 0.002 | 0 | 22.38 | 0 | 0.055 |  |  |
| Socialness | -0.003 | 0.001 | -3.6 | 0 | 0.001 |  |  | 0 | 0 | -0.31 | 0.754 | 0 |  |  |
| Concreteness | -0.003 | 0.001 | -2.04 | 0.041 | 0 |  |  | 0 | 0 | 1.46 | 0.145 | 0 |  |  |
| Valence Extremity | -0.009 | 0.001 | -6.09 | 0 | 0.003 |  |  | -0.001 | 0 | -3.52 | 0 | 0.001 |  |  |
| Semantic Diversity | -0.025 | 0.004 | -6.01 | 0 | 0.003 |  |  | -0.004 | 0.001 | -5.89 | 0 | 0.004 |  |  |

*Note.* b represents unstandardized regression weights. SE represents the standard error of the regression weights. sr2 represents the semi-partial correlation squared. zRTs standardized reaction times. *p < .05;* ***p < .01;*** p < .001

There were 7,010 items for which we had values for all variables of interest in the analysis predicting performance in the ECP word knowledge task. Descriptive statistics and zero-order correlations between all variables of interest from this dataset are reported in supplementary Table S2. The statistical results are reported in Table 4 and the standardized coefficients and illustrated in Figure 3B. In this analysis, the control variables were all significant predictors of response latencies – RTs were faster for words that are shorter, more frequent and acquired earlier. There was significant improvement in model fit with the addition of the semantic variables, which accounted for a further 0.78% of variance in recognition RTs. All semantic variables were significant predictors, with faster RTs for words with increased socialness, concreteness and valence extremity and for those encountered in more semantically diverse contexts. The control variables were all significant predictors of the proportion of participants reporting not knowing a word, with words that are longer, more frequent and acquired earlier being more prevalent. There was significant improvement in model fit with the inclusion of the semantic variables, which accounted for an additional 0.83% of variance in in ECP proportion unknown. Valence and semantic diversity were the only significant semantic predictors – words that are more valenced and encountered in more semantically diverse contexts were reported as known by more participants.

